AL-TP-1991-0018

AD-A240 566

A R M S T R O N G

L A B O R A T O R Y

COGNITIVE FACTORS INVOLVED IN THE FIRST STAGE OF PROGRAMMING SKILL ACQUISITION

Carmen M. Pena
William C. Tirre

HUMAN RESOURCES DIRECTORATE
MANPOWER AND PERSONNEL RESEARCH DIVISION
Brooks Air Force Base, TX 78235-5000

July 1991

Interim Technical Paper for Period March 1989 - March 1991

Approved for public release; distribution is unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5000

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

William C. Tirre
WILLIAM C. TIRRE
Project Scientist

William E. Alley
WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Research Division

R. W. Alford
ROGER W. ALFORD, Lt Col, USAF
Chief, Manpower and Personnel Research Division

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1991	Interim – March 1989 – March 1991

4. TITLE AND SUBTITLE
Cognitive Factors Involved in the First Stage of Programming Skill Acquisition

5. FUNDING NUMBERS
PE – 61102F
PR – 2313
TA – T1
WU – 33

6. AUTHOR(S)
Carmen M. Pena
William C. Tirre

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Armstrong Laboratory
Human Resources Directorate
Manpower and Personnel Research Division
Brooks Air Force Base, TX 78235-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
AL-TP-1991-0018

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This study examined the role of cognitive abilities in the first stage of programming skill acquisition. Three hundred and five Air Force recruits were given 1.5 hours of computer-based instruction on PASCAL. They were also administered computer-based tests assessing working memory, general reasoning, programming related knowledge, and component skills underlying algebra word problem solving. Additional test scores reflecting verbal knowledge were available from personnel records. Structural equation modeling was used to test alternative models of PASCAL learning, general cognitive abilities, and algebra word problem solving. The factors resulting from these analyses were then interrelated using structural models. The results showed that a model of PASCAL learning reflecting declarative and procedural factors was not superior in fit to a single factor model. In the general cognitive abilities domain, a three-factor model positing separate working memory, reasoning, and verbal knowledge abilities provided a better fit to the data than a single factor (g) model. Three factors were found to underlie algebra word problem solution. In interrelating the three sets of factors, we found that word problem translation added to significant contributions from more general abilities such as general reasoning, working memory, and verbal knowledge in predicting success in programming skill acquisition.

14. SUBJECT TERMS
cognition
cognitive ability
computerized testing
individual differences
learning
learning ability

15. NUMBER OF PAGES
32

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

Carmen A. Pena conducted this research to fulfill the requirements for the Master of Science degree in psychology at St. Mary's University. This study was conducted as part of the Learning Abilities Measurement Program (LAMP), a basic research program cosponsored by the Air Force Human Resources Laboratory (AFHRL) and the Air Force Office of Scientific Research (AFOSR). The goals of this program are to specify the basic parameters of learning ability, to develop measurement techniques for cognitive abilities, and to explore the feasibility of a model-based system of cognitive measurement.

The authors wish to thank Dr. Patrick Kyllonen for his guidance in the structural equation modeling and for his development of the original version of the PASCAL tutorial. We also thank Dr. Valerie Shute and Dr. William Alley for their comments on an earlier version of this paper. Thanks are also due to Mr. Rich Walker, Ms. JoAnn Hall, and Mr. Jim Bloch of the OAO Corporation who provided programming support, and to Mr. Roy Chollman and his staff at the AFHRL Experimental Testing Facility at Lackland AFB, who collected the data.

COGNITIVE FACTORS INVOLVED IN THE FIRST STAGE OF PROGRAMMING SKILL ACQUISITION

SUMMARY

In this study we examined the role of various cognitive abilities in the first stage of programming skill acquisition. A sample of U.S. Air Force recruits was given computer-based instruction on PASCAL. They were also administered a battery of computer-based tests assessing working memory, general reasoning skill, programming-related knowledge, and component skills underlying algebra work problem solving. Additional test scores reflecting general verbal knowledge were available from personnel records. We found that algebra word problem translation skills added to more general abilities such as general reasoning, working memory, and verbal knowledge in predicting success in the first instructional encounter with PASCAL programming. The contribution of word problem translation skills was independent of these other abilities and of previous formal instruction in another programming language. This study suggests that prediction of success in a technical specialty might be augmented with the addition of tests that assess specific cognitive skills similar to those required by the technical specialty. The component skills required in algebra word problem solving (e.g., problem type identifcation, decomposition and sequencing, and problem translation) are similar to those required in PASCAL programming. Cognitive task analyses of technical specialties may prove beneficial in identifying cognitive skills not currently assessed by the standard test battery, the Armed Services Vocational Aptitude Battery.

INTRODUCTION

Background

The purpose of this study was to identify the cognitive abilities that enable some students to learn programming skills faster than others. The specific focus of this research was on the initial stage of skill acquisition: the student's first encounter with a new programming language.

Studies of individual differences in skill acquisition have shown that working memory is especially important in the initial phase of learning where the student is acquiring the declarative knowledge, i.e., the facts and concepts needed for skill execution (Woltz, 1988; Kyllonen & Woltz, 1989). Skill execution in this phase is a controlled process demanding attentional resources or working memory (Anderson, 1982, 1987; Ackerman, 1987, 1988). When working memory fails, the student loses information needed for task performance which may lead to false starts and faulty solution paths. With further practice the student detects and eliminates misconceptions about task performance and transforms the declarative knowledge into a more easily executable procedural representation. At this point, the role of working memory becomes less important (Woltz, 1988).

Ackerman (1987, 1988) has proposed that individual differences in the declarative stage are primarily due to differences in general cognitive ability (g) which he equated with attentional resource capacity (another term for working memory capacity) and to content-related abilities (such as spatial ability when the skill requires processing spatial information). Ackerman's general cognitive ability measure combines several declarative and procedural knowledge (cognitive skill) subtests from the Armed Services Vocational Aptitude Battery (ASVAB) into one score. Disassembling this composite score would allow a finer-grained analysis of how potentially distinguishable cognitive factors predict learning.

Declarative knowledge is one ingredient of Ackerman's g. Knowledge facilitates comprehension and retention of new information by making concepts in semantic memory more accessible, by making the activation of semantic relations more automatic and less effortful, and by facilitating the use of elaborative strategies (Bjorklund, 1987). Studies of verbal learning have shown that students with more knowledge learn new associations more quickly (Kyllonen & Tirre, 1988; Kyllonen, Tirre, & Christal, 1989). This effect has been found to be independent of other cognitive abilities such as inferential ability and working memory (Tirre, 1989). The knowledge effect generalizes to prose learning as well (for review see Anderson & Pearson, 1984).

A second ingredient of Ackerman's g is reasoning ability, reflected primarily in the arithmetic reasoning, math knowledge, and mechanical comprehension subtests on the ASVAB. Kyllonen and Christal (1991) have found substantial overlap between working memory and reasoning factors to the extent that they suggest that reasoning tests measure little more than working memory. Their conclusions are warranted given the types of working memory and reasoning tests they administered. A different conclusion might be obtained with a different selection of tests. Thus, it would be instructive to see if working memory and reasoning made unique contributions to the explanation of skill acquisition.

Several psychometric studies have found substantial correlations between reasoning and programming criteria (for review see Pena, 1989). These studies suggest that reasoning is related to programming skill, but they do not address the issue of what specific aspects of reasoning ability are involved in learning to program.

Mayer, Dyck, and Vilberg (1986) administered a battery of specifically focused reasoning skills to a sample of students taking a course in BASIC. The battery of tests measured the following abilities: (1) word problem translation, (2) word problem solution, (3) following procedures, (4) following directions, (5) logical reasoning, (6) visualization ability (paper folding task), (7) verbal ability (decoding of verbal messages), and (8) arithmetic computation (solving addition and division problems). Significant correlations ($p < .05$) between BASIC exam scores and scores on the cognitive ability tests were obtained for word problem translation ($r = .55$), word problem solution ($r = .56$), following procedures ($r = .44$), following directions ($r = .44$), logical reasoning ($r = .29$), and visualization ability ($r = .31$). Based on a stepwise multiple regression analysis, Mayer et al. (1986) found that (1) word problem translation skills, (2) word problem solution skills, and (3) following directions skills were better predictors of success in learning BASIC than general intellectual ability as measured by the logical reasoning and verbal ability tests. Scores on the word problem translation, word problem solution, and following directions tests together accounted for 50 percent of the variance in BASIC exam scores.

Mayer (1985; 1987; Mayer et al., 1986) has argued that learning to program involves acquiring both syntactic and semantic (conceptual) knowledge. Syntactic knowledge in programming includes specific terminology (e.g., knowing the meanings of keywords such as READ, DATA, INPUT) and format rules for writing code. For example, a loop has the following format in BASIC:

```
FOR variable = x TO y [STEP z]
    .
    .
NEXT [variable] [,variable]...
```

Conceptual knowledge, on the other hand, consists of the concepts and structures that underlie the surface form of code. As such, it is relatively independent of any particular programming language. Mayer believes that becoming an expert programmer means acquiring accurate conceptual knowledge. Sometimes a student's existing conceptual knowledge will transfer to a new domain. Thus, students who performed well on the problem translation, following directions, and word problem solution tests also learned BASIC programming well in the Mayer et al. (1986) study because they were able to apply conceptual knowledge to the programming domain. Mayer believes that this conceptual knowledge can be taught directly. Bayman (1983; cited in Mayer, 1987) has found that instruction that represents or models the transactions the computer goes through in executing a line of code significantly enhances learning over the standard form of instruction that emphasizes syntax.

Hypotheses for the Present Study

Hypotheses were suggested by a theoretical framework based on the theory of Cattell (1971). This framework proposes that working memory capacity and basic reasoning mechanisms are early-determined cognitive characteristics of humans and as such, have the highest causal priority. The framework also asserts that verbal knowledge is acquired through the investment of these cognitive abilities in school and other learning experiences, and that this verbal knowledge facilitates the acquisition of more specific knowledge in a particular domain, such as mathematics or programming. The latter hypothesis is identical to the knowledge hypothesis of Anderson and Freebody (1981) which proposed that vocabulary predicts comprehension because it reflects the amount of relevant cultural knowledge represented in the student's schemata.

Following Mayer et al. (1986) and a cognitive task analysis by Brooks (1977), we hypothesized that specific reasoning skills, i.e., problem type identification, problem decomposition and sequencing, and problem translation in algebra word problem solving, would add to general reasoning, working memory, and knowledge in predicting learning. We hypothesized that the type of processing involved in solving algebra word problems was approximately isomorphic with that involved in programming a computer. Skill in the components of word problem solution should transfer to the programming domain.

In addition, we hypothesized that domain specific knowledge would make a significant unique contribution to prediction. This hypothesis is identical to the instrumentalist hypothesis of Anderson and Freebody (1981) which suggests that the effects of knowledge on comprehension are rather specific. For example, a student who already knows what a variable is in mathematics is more likely to understand a discussion of variables in PASCAL programming.

In our study we correlated measures of working memory, general reasoning skill (referred to as g_f, or fluid intelligence by Cattell, 1971), specific reasoning components selected on the basis of previous cognitive task analyses (e.g., the Mayer et al., 1986 and Brooks, 1977 studies), general verbal knowledge scores, and specific prior knowledge of programming concepts with learning outcomes obtained from a computer-based tutorial on PASCAL programming. The goals of this study were (1) to test competing structural models concerning the interrelationships of the cognitive variables proposed as predictors and indicators of skill attainment; and (2) to test hypotheses concerning the unique contributions of each variable set in predicting early programming skill acquisition.

METHOD

Subjects

The subjects were 305 military recruits in their 11th day of basic training at Lackland Air Force Base, Texas. The number of subjects that were usable varied with the analysis performed. Fifty-eight subjects were eliminated because they had taken a course in PASCAL programming. Of the remaining 247 subjects, 117 were Air National Guard or Air Force Reserve recruits who did not have the ASVAB subtest scores needed for some analyses. Thus, in the analyses which included ASVAB variables, the sample size was 130. In the Algebra Problem Solving Test or the PASCAL tutorial analyses, the sample size was 247.

The intelligence level of the subject sample was slightly above average in terms of national norms (60th percentile) on the ASVAB. Most of the subjects possessed at least a high school education; their average age was 20 years.

Apparatus

All of the tests, except for the three ASVAB subtests were computer-administered on Zenith Z-248 (IBM AT-compatible) computers. Z-248s were equipped with enhanced graphics adaptor cards, color monitors, 20 megabyte hard disks, and standard keyboards.

Criterion Task: PASCAL Tutorial

The PASCAL learning task was a self-paced tutorial designed to teach introductory PASCAL programming concepts and to assess comprehension of these concepts (Kyllonen & Soule, 1988). It was made up of 10 different sections consisting of 6 testing sections and 4 instruction sections. Subjects were initially given declarative information identifying what a program is, how it is executed, and what the PASCAL symbols used to designate the four arithmetic operators (+, -, *, and /) mean; variables and constants were explained is well. The second instruction block illustrated how the computer processes a program step-by-step, displaying the results of each processing step. Subjects were also taught the concept of loops in this section. In the third instruction block subjects learned how to write a program using a loop that computes and prints out the powers of 2 that are less than 10. The fourth instruction section contained more elaborate explanations of loops as well as examples of the different errors that a program can contain, such as an infinite loop.

Each instruction section was followed by a comprehension test. The first four tests were true/false declarative knowledge tests. The last two tests were more procedural; subjects were required to do such things as detect the errors in short PASCAL programs and fill in the missing parts of a program.

Working Memory Tests

Alphabet Recoding Task (WMAR)

The alphabet recoding task (Woltz, 1988) was designed to measure working memory capacity by requiring subjects to store and process information simultaneously. Subjects were shown from one to three letters, one at a time. Each letter was displayed for 1 second and then removed from the screen. The letters were followed by one of the following numbers: -3, -2, -1, +1, +2, +3, which indicated how many letters the subject must move backward (minus) or forward (plus) to locate the new letters. Subjects were instructed to recode the letters according to the number displayed. For example, if the letters C, G, and K were followed by a +2 the correct answer would be E, I, and M. Because the letters were flashed on the screen one at a time for 1 second each, subjects had to maintain them in memory while they attempted to recode them; thus. subjects had the dual tasks of storing and processing information simultaneously. There were 39 items in the task. Solution time and percent correct scores were recorded for this task.

Backward Digit Span Task (WMBDS)

In the backward digit span task, subjects were presented with lists of numbers ranging from three to eight digits long. Each digit was presented separately on the screen for .75 seconds. Subjects were instructed to memorize the numbers and then recall them in reverse order. Thus, the subjects must store the string in active memory while carrying out a mental transformation. Subjects had an unlimited time to respond. Accuracy and response time feedback were given for all 24 items. Percent correct scores were recorded for this task.

Continuous Paired Associates Task (WMCPA)

In the continuous paired associates task (Hunt & Lansman, 1982), subjects must maintain the current value for a variable list of letters. Subjects were presented with 2, 3, 4, 5, or 7 letter-digit pairs. A letter-digit pair was displayed for 3 seconds. Subjects were then required to recall the number paired with a particular letter displayed on the screen. After each question, subjects were given new letter-digit pairs to learn which involved the same letter presented on the previous question, but paired with a new number. When queried with a letter, such as A = ? subjects were expected to respond with the number which was last paired with that letter. Subjects were given right/wrong feedback and were given only 10 seconds to respond. In this task subjects stored and processed information simultaneously—they had to maintain the old letter-digit pairs in working memory while attempting to encode the new ones. Information recall typically decreased as memory load increased.

General and Domain Specific Reasoning Tests

Culture Fair Intelligence Test (CFIT)

The CFIT (Scale 3) is a speeded, nonverbal, general measure of intelligence (Cattell, 1973). The test items consist of abstract symbols and drawings designed to minimize verbal, cultural, and educational influences. There are four subtests to Scale 3 of the CFIT: Series, Classification, Matrices, and Conditions. Several practice items precede each subtest. The Series subtest consists of items in which three progressive pictures are arranged in a row followed by a blank, and a set of six alternatives. The task is to choose the alternative that best completes the series. The Classification subtest consists of a set of pictures in which the task is to choose the two pictures that do not belong in the set of five pictures. The Matrices subtest consists of a sequence of four to nine boxes arranged in a matrix with a blank box in the bottom right corner. A set of six alternative boxes is presented to the right of each matrix, and the task is to choose the one that best completes the sequence. The Conditions subtest consists of items in which a single picture is used to prescribe a set of conditions. For instance, a dot might be placed in a circle but outside of a box which is also within the circle. Examinees must choose the picture that best meets the conditions prescribed by the stimulus. Performance was indicated by percent correct scores.

Algebra Word Problem Solving Test (WPS)

The algebra word problem solving test assessed three different problem solving abilities in algebra which mirror the stages in Brooks' (1977) model of programming behavior. The test was composed of three sections: (1) word problem identification (PI), (2) word problem decomposition and sequencing (DS), and (3) word problem translation (PT). Each section contains 20 items. In Section 1, subjects were presented with different types of word problems and were asked to identify the type such as area, mixture, triangle, and probability. For instance, subjects might be presented with the following progressions problem:

> A tennis ball, dropped 128 ft., rebounds on each bounce 1/2 of the distance from which it fell. How high does it bounce on its 9th rebound?

Subjects had a total of 12 different problem types from which to select. The problem types included triangle, distance/rate/time, averages, scale conversion, ratios, interest, area, mixture, probability, number, work, and progressions.

In Section 2, subjects were again presented with word problems. Their task was to break the problem down into the different mathematical operations (add, subtract, multiply, or divide) needed to solve the problem and to put them into the correct linear order. An example of such a problem is:

> Palmer drove out to the country and back home. He drove out at 45 m.p.h. If he drove 21 m.p.h. slower on his way home, what was his average rate roundtrip?
>
> 1) Subtract, Multiply, Add
> 2) Subtract, Add, Divide
> 3) Divide, Subtract, Add
> 4) Add, Multiply, Subtract

To obtain the correct solution to this problem, subjects would have to first subtract 21 from 45, add this quantity to 45, and divide by 2, so the correct answer is selection "2".

The third section of the test required subjects to translate verbal problem statements into mathematical equations. One possible problem might be:

> Mr. Smiley bought 4 cans of chicken soup and 5 cans of tomato soup for a total cost of $2.70. Each can cost the same amount. How much did each can cost?
>
> 1) $4n = 2.70 / 5x$
> 2) $4n / 5n = 2.70$
> 3) $4n \times 5n = 2.70$
> 4) $4n + 5n = 2.70$

The correct answer for this problem is selection "4". The entire test took approximately 1 hour to complete. Subjects were given accuracy and response time feedback on each section of the test.

General (ASVAB) and Specific Knowledge Tests

The ASVAB is a group administered paper-and-pencil test used for selection and placement purposes in the military. The ASVAB tests the "developed abilities" of individuals between the ages of 16 and 23 at a high school level (U.S. DOD, 1984). It consists of 10 separately timed subtests (8 power and 2 speeded). Only three subtests were used in this study as measures of general knowledge, namely: General Science knowledge, Paragraph Comprehension, and Word Knowledge. All of the items in each of these three subtests are in multiple-choice with four alternatives.

General Science (GS)

The GS subtest measures knowledge of science covered by texts for junior and senior high school science courses such as biology, earth science (geology, astronomy, and meteorology), and chemistry. It contains 25 items, 45% life science questions, 45% physical science items, and 10% earth science items.

Paragraph Comprehension (PC)

The PC subtest is an assessment of an individual's ability to understand short paragraphs. The 15 questions of the PC subtest can be broken down into four categories: memory for literal detail, recognition of the main idea, making inferences and generalizations, and recognizing and understanding sequential, causal, and comparative relationships.

Word Knowledge (WK)

The WK subtest is a standard four-choice test of vocabulary. It contains 35 items; approximately 35% of the words are nouns, 30% verbs, and 35% adjectives.

Pretest of Computer Programming Knowledge (SPKN)

This pretest was designed by Shute and Kyllonen (1990) to assess an individual's prior knowledge of programming and related concepts. Part 1 was a questionnaire concerning years of education, number of programming courses, familiarity with computers and PASCAL, and number of math and English courses. Part 2 contained 39 questions regarding such programming concepts as: sums, expressions, integers, products, real numbers, assignment statements, strings, constants, data, and variables. Part 2 also contained 20 additional true/false questions which assessed knowledge of the 10 concepts. The final section consisted of five programming problems in which subjects were shown a PASCAL program and asked to indicate whether or not it contains an error and, if so, what kind of error. Percent correct scores were recorded on Sections 2 and 3 of the test.

Procedure

Subjects were tested in groups of 30-35 in 3-hour testing/learning sessions. Prior to testing, subjects were given a general briefing explaining the purpose of the various experimental tasks. They were also required to complete familiarization exercises to improve keyboard skills and expedite response entry. Subjects first took the Pretest of Computer Programming Background and Knowledge, then they worked through the learning/testing sections of the PASCAL tutorial which lasted approximately 1.5 hours. After a five-minute break, the order of administration of the rest of the tests was randomized for each subject.

RESULTS

The data were analyzed in two ways. In the first set of analyses we sought to test alternative structural models of both the predictor and criterion variable sets. For the general cognitive abilities, we asked whether it was reasonable to distinguish separate general reasoning, working memory, and verbal knowledge factors. For the domain-specific (algebra) reasoning test, we tested the hypothesis that three reasoning abilities were being assessed. Likewise, on the criterion side, we tested the hypothesis that separate declarative and procedural factors could be distinguished.

In the second set of analyses we adopted an exploratory mode in attempting to integrate the models for the general cognitive abilities, domain-specific (algebra) reasoning abilities, and the PASCAL tutorial. The two predictor domains were first interrelated, and then all predictors were correlated with the criterion factor.

Declarative Understanding versus Procedural Competence

Table 1 displays the correlations between the predictors and PASCAL tutorial performance and its two subcomponents. The first, which we labeled declarative understanding, consisted of the four tests which assessed comprehension with true/false items. The second, labeled procedural competence, consisted of the two tests in which the student demonstrated some procedural skill by detecting errors in PASCAL programs and by filling in blanks in PASCAL code.

TABLE 1. CORRELATIONS BETWEEN PERFORMANCE ON THE PASCAL TUTORIAL AND COGNITIVE ABILITY MEASURES

	r with		
Cognitive Ability	Total	Declarative Understanding	Procedural Competence
General Verbal Knowledge	.44	.36	.47
General Science	.43	.39	.39
Word Knowledge	.24	.16	.31
Paragraph Comprehension	.28	.21	.35
Specific Computer Knowledge	.45	.39	.42
Mathematical Concepts	.36	.32	.33
Programming Concepts	.31	.31	.23
General Reasoning Skill	.53	.50	.44
CFIT1: Series	.45	.47	.30
CFIT2: Classifications	.26	.23	.25
CFIT3: Matrices	.31	.26	.31
CFIT4: Conditions	.40	.38	.32
Algebra Word Problem Solving	.56	.53	.46
Problem Identification	.39	.36	.35
Problem Decomposition	.42	.40	.34
Problem Translation	.57	.56	.44
Working Memory Capacity	.51	.52	.37
Continuous Paired Assoc	.48	.47	.35
Alphabet Recoding	.50	.48	.38
Backward Digit Span	.29	.31	.17

Note. N = 130. Critical value of r (p = .025, 1-tailed) is .15.

Composite scores were computed for verbal knowledge, working memory, and general reasoning; but the domain-specific tests, i.e., programming/math concepts and the three algebra word problem solving components tests were retained as individual scores (see Appendix for means, standard deviations, and reliability estimates).[1]

Correlations of these composite scores with PASCAL learning were significant and ranged from .44 to .56. The highest correlation (.56) was with the word problem solving score, as expected from previous research.[2] There was no evidence of the cognitive abilities correlating differently with declarative understanding and procedural competence. This finding might be expected given the substantial overlap between these two outcome measures, r = .59.

To determine what kind of outcome measure should be used in subsequent analyses, we compared the relative fits of three different factor models using the EQS structural equation modeling program (Bentler, 1989). Model 1 (Fig. 1) was a one-factor model making no distinction between declarative and procedural subtests. Models 2 and 3 were two-factor models (Figs. 2, 3). Model 2 proposed two correlated factors and Model 3 proposed two orthogonal factors; i.e., that Factor 1 entered into all subtests and Factor 2 entered into only the procedural subtests. There were four declarative tests and two procedural tests administered in the tutorial. The latter two tests were split into odd- and even-numbered scores to increase the number of indicators up to four per factor.

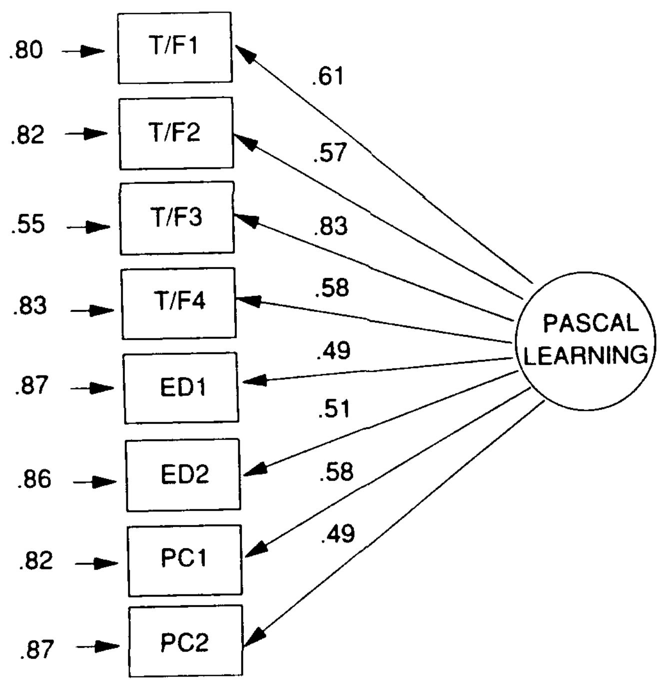

Figure 1. Single factor model (Model 1) of performance on the PASCAL tutorial. T/F = true-false test, ED = error detection test, and PC = program completion test.

[1] The programming knowledge pretest which presented PASCAL code with errors to detect was unreliable and was discarded.

[2] Shute and Pena (1990) found similar correlations for a paper-and-pencil version of the Algebra Word Problem Solving Test. Total score on this test correlated .69 with learning outcome on a PASCAL intelligent tutor (N = 257).

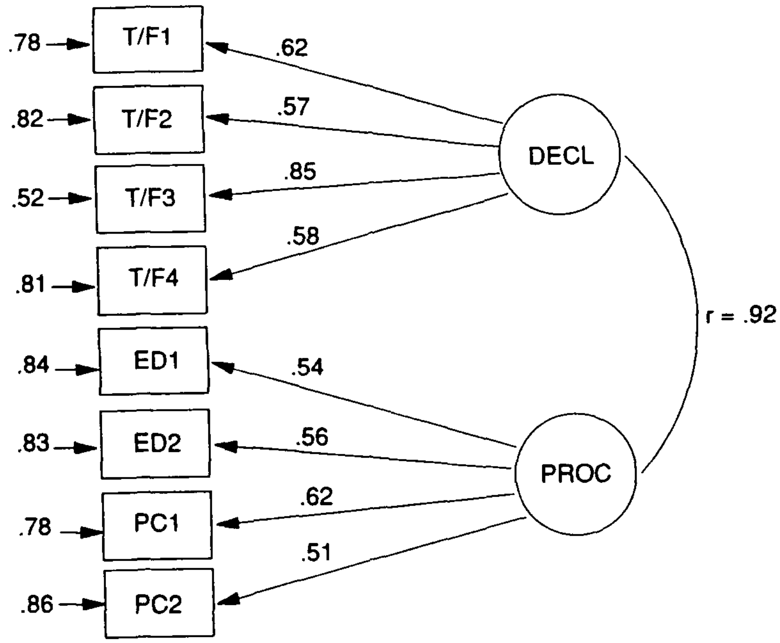

Figure 2. Oblique two-factor model (Model 2) of performance on the PASCAL tutorial. T/F = true-false test, ED = error detection test, and PC = program completion test.

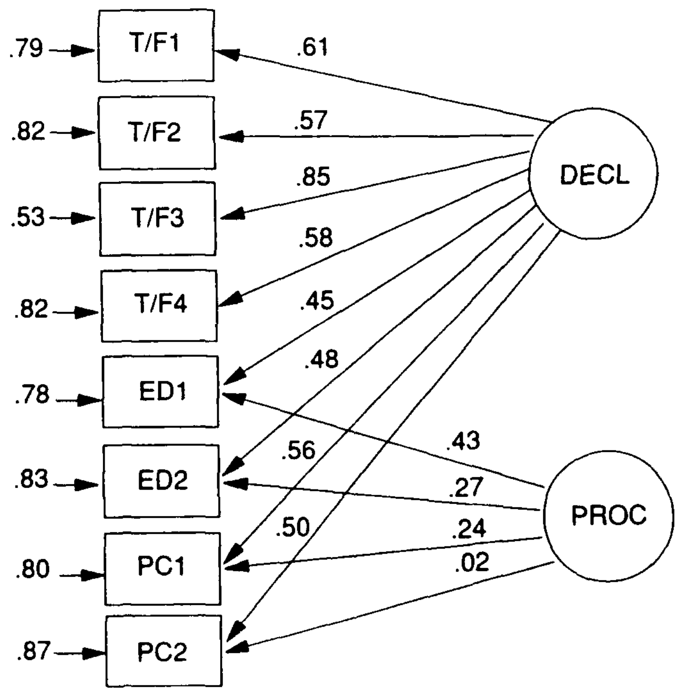

Figure 3. Orthogonalized two-factor model (Model 3) of performance on the PASCAL tutorial. T/F = true-false test, ED = error detection test, and PC = program completion test.

Confirmatory factor analysis showed that each of the factor models fit the data well. Each model resulted in nonsignificant chi squares (p = .227, .298, .467, for Models 1-3, respectively) and high goodness-of-fit indices (Bentler-Bonett non-normed fit = .987, .991, and 1.000 for Models 1-3, respectively). The one-factor model was embedded in the orthogonalized two-factor model, thus permitting a *chi-square* difference test to see if there was a significant decrease in fit for the one-factor model. The difference chi-square of 8.56 with 4 degrees of freedom was not significant (critical value for p = .05 is 9.488). Thus, there is no appreciable loss in fit with the one-factor model. Also, because the oblique two-factor solution indicated a factor intercorrelation of .92, we decided to accept the one-factor model. We believe it is reasonable to assume that most subjects were in the declarative stage of skill acquisition after only 1.5 hours of instruction and that knowledge was not yet becoming compiled, i.e., transformed into a procedural representation.

Analysis of General Cognitive Abilities

Do working memory capacity, general reasoning skill, and verbal knowledge simply reflect the same underlying ability, viz., g? Or is it reasonable to hypothesize that three distinguishable abilities underlie performance on these types of tests? We examined two highly related models 4a and 4b (Figs 4a and 4b) as alternatives to a pure g model.

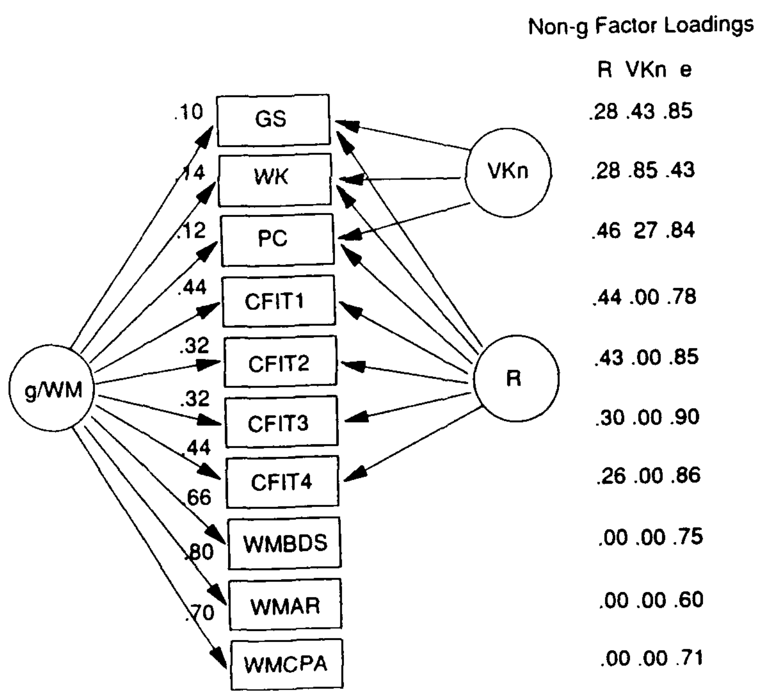

Figure 4a. Orthogonalized structure model (Model 4a) of working memory (WM), verbal knowledge (VKn), and reasoning skill (R). GS = General Science, WK = Word Knowledge, PC = Paragraph Comprehension, CFIT = Culture Fair Intelligence Test (Parts 1-4), WMBDS = backward digit span, WMAR = alphabet recoding task, WMCPA = continuous paired associates task.

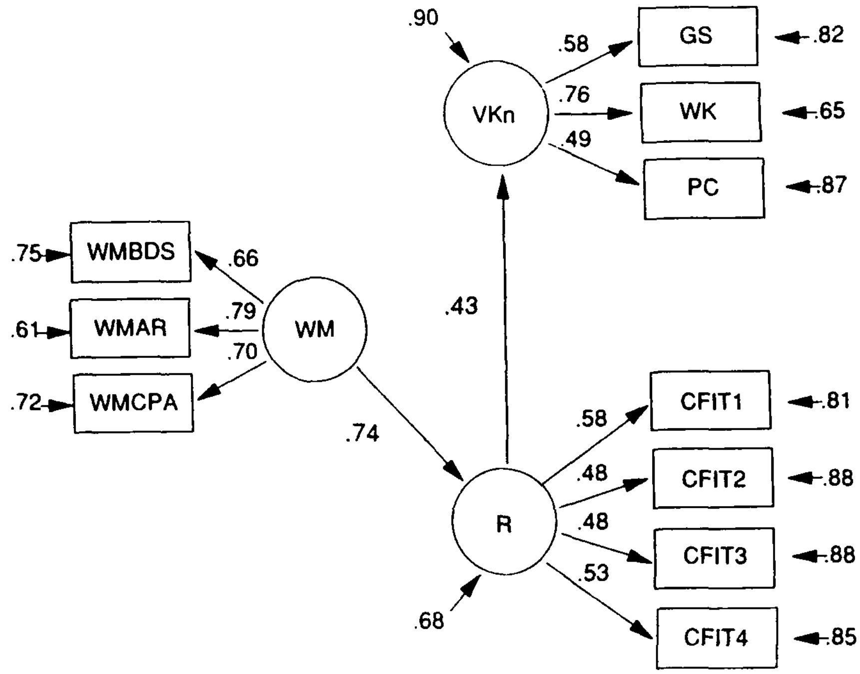

Figure 4b. Causal model (Model 4b) of working memory, reasoning skill, and verbal knowledge.

Model 4a is a purely structural model — a "snapshot" of the current status of these cognitive abilities. According to this model, performance on working memory tests, reasoning tests, and knowledge tests is enabled by g, which we equate with working memory capacity. The model also asserts that performance on reasoning and knowledge tests is enabled by a reasoning ability that is orthogonal to g/working memory. This part of the model is consistent with the notion that knowledge is acquired through reasoning processes such as induction and deduction. The model also proposes that knowledge test scores are affected by a third factor orthogonal to g/working memory and reasoning. This third factor probably reflects the amount of knowledge stored in long-term memory, and it might indirectly reflect quality of educational experiences.

Model 4b is a developmental model that reflects the hypotheses of our theory inspired by Cattell's (1971) investment theory. Here we equate working memory with Cattell's g_f, which enables the individual to acquire reasoning skills, which in turn, enables knowledge acquisition.

Confirmatory factor analyses indicated a good fit of the Models 4a and 4b to the data. Model 4a resulted in a chi-square of 19.80 (p = .757, df = 25) and Bentler-Bonett non-normed fit index of 1.006. The loadings of the knowledge tests on g/working memory were nonsignificant, suggesting a minimal involvement of working memory in responding to knowledge test items. The *one-factor model* (Model 4c) was also tested, and did not fit the data as well (*chi-square* = 54.66, p = .018, df = 35, non-normed fit index = .965). The difference *chi-square* (34.86, df = 10) was significant, p < .001, indicating a superior fit for the three-factor model.

Model 4b resulted in a chi-square of 28.36 (p = .697, df = 33) and a non-normed fit index of 1.004. Working memory was found to be a strong determinant of reasoning skill (path coefficient = .74) and reasoning skill a moderately strong determinant of verbal knowledge (path coefficient = .43). Together, these analyses indicate it is reasonable to examine working memory, reasoning skill, and verbal knowledge as separate cognitive factors.

Analysis of the Algebra Word Problem Solving Test

Confirmatory factor analyses were conducted on the word problem solving test to test our hypotheses concerning the processing components underlying performance on this task. We hypothesized that word problem translation was the most psychologically complex task in the word problem solving test. To test this hypothesis directly we set up three confirmatory factor analysis models again using the EQS structural equation modeling software (Bentler, 1989). The three word problem solving subtests were split into separate scores for Parts 1 and 2 in order to achieve an identified model. However, this action resulted in a condition known as *empirical under-identification* (see Kenny, 1979; Rindskopf, 1984; cited in Bentler, 1989). Consequently, we further subdivided the two parts of the problem translation task into two additional portions consisting of odd- and even-numbered items. This division gave us additional indicators and the empirical under-identification problem was solved.

Our first model (Model 5; see Figure 5) hypothesized three underlying components: a problem identification ability, a problem decomposition and sequencing ability, and a problem translation ability. The problem identification test was assumed to be the least inclusive and to assess only the problem identification component; the decomposition test was assumed to measure both problem identification and decomposition; and the problem translation test was assumed to assess *all* three underlying components.

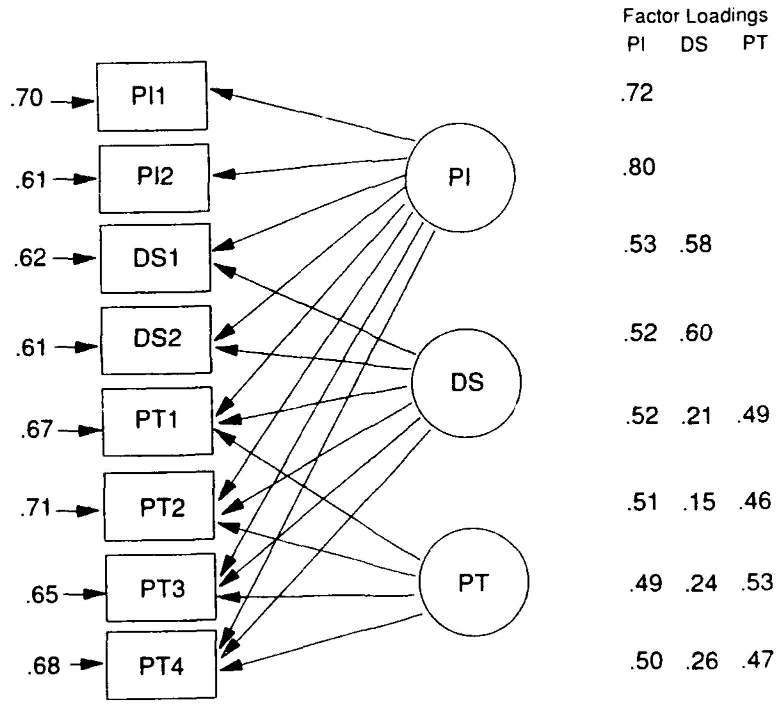

Figure 5. Three-factor model (Model 5) of the algebra word problem solving test. PI = problem identification skill, DS = decomposition and sequencing skill, and PT = problem translation skill.

Contrast this design with Model 6 (Fig. 6) which proposed that two correlated components were being assessed: a simple problem identification ability which affected performance in the problem identification test and a more complicated problem solution ability which affected performance in both the problem decomposition and translation tests. A second plausible model, Model 7 (Fig. 7), asserted that only one word problem solving ability was assessed by each of the three tests.

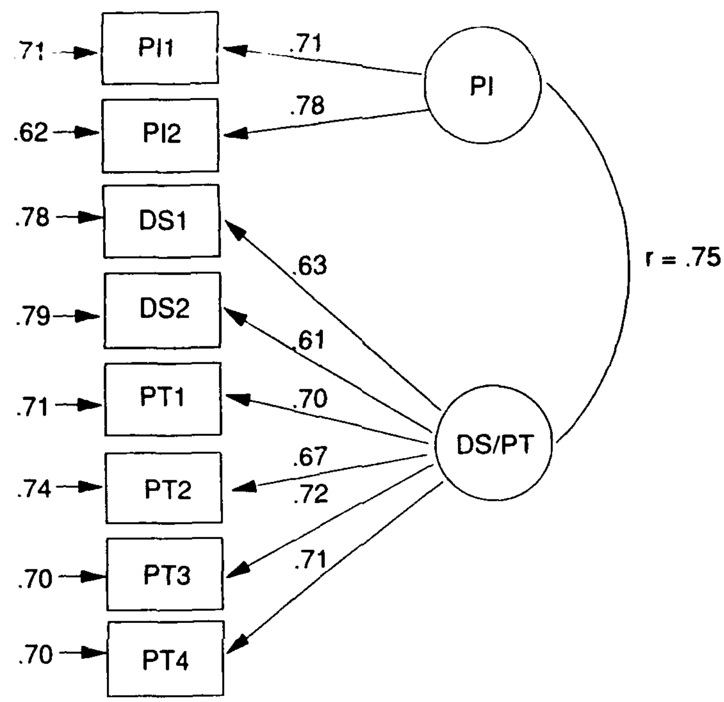

Figure 6. Two-factor model (Model 6) of the algebra word problem solving test. PI = problem identification skill, DS/PT = problem decomposition, sequencing, and translation skill.

The EQS goodness-of-fit results (Table 2) indicated support for Model 5 over Models 6 and 7. Model 5 resulted in a nonsignificant chi-square. In contrast, both Models 6 and 7, had significant chi-square values, and can be rejected.

In sum, these structural equation analyses indicated that three latent processing abilities underlie performance on the algebra word problem test and that the problem translation subtest is the most complex psychologically, involving all three latent abilities.

Interrelating the General and Specific Cognitive Abilities

We have reported the results of our confirmatory factor analyses of three sets of measures, i.e., the PASCAL tutorial learning indicators, the general cognitive abilities, and the specific algebra word problem solving components. To interrelate the three sets of factors, we adopted an exploratory mode of analysis. The analyses that follow are exploratory in that we generally tested the full model (i.e., each possible factor was hypothesized to be a predictor) and allowed EQS to identify the reduced model.

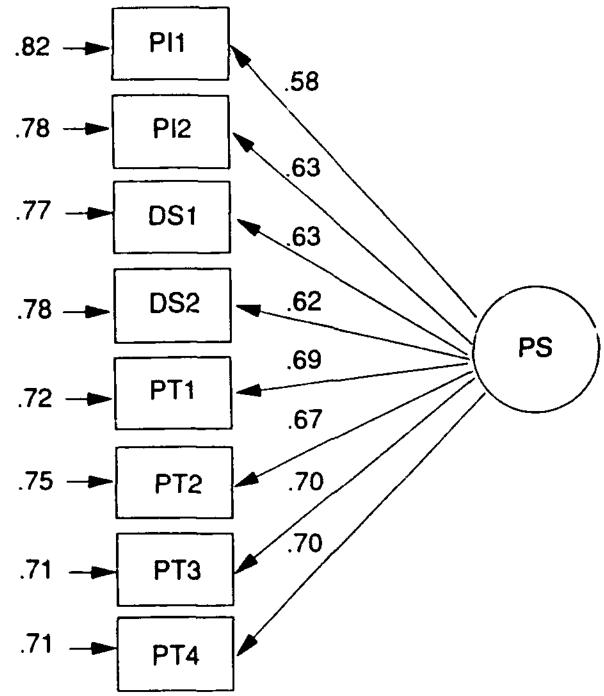

Figure 7. *One-factor model (Model 7) of the algebra word problem solving test.* PS = word problem solving skill.

TABLE 2. GOODNESS-OF-FIT STATISTICS FOR CONFIRMATORY FACTOR ANALYSES OF THE WORD PROBLEM SOLVING TEST

Model	Mean Off-Diagonal Residual	χ^2	*df*	*p*	*NFI*	*NNFI*	*CFI*
5 3 Factors-O	.0191	17.35	10	.067	.995	.994	.998
6 2 Factors-C	.0419	65.20	19	.001	.905	.897	.930
7 1 Factor	.0521	92.67	20	.001	.865	.846	.890

Note. O and C signify orthogonal and correlated factors, respectively. NFI and NNFI are the Bentler-Bonett normed and non-normed fit indices, respectively. CFI stands for the comparative fit index.

Guided by our theory, we assessed the fit of a model that proposed g/working memory, reasoning, and verbal knowledge as determinants of the specific problem solving skills. Factor loadings found for previous models (i.e., 4a and 5) were input as fixed values. As it turns out, problem translation was not significantly predicted by any of the three general cognitive abilities. In contrast, problem type identification was predicted well (69.6% of the variance) by g/working memory (path coefficient [*beta*] = .75, z = 8.62) and verbal knowledge (*beta* =

.37, z = 4.08). Problem decomposition and sequencing was predicted with modest precision (12.4%) by the reasoning factor (*beta* = .35, z = 2.18). The final model which eliminated all nonsignificant paths between factor sets resulted in a *chi-square* of 183.79 (*df* = 163, p = .127) and a non-normed fit index of .996. Thus, the data appear to support a model that portrays problem translation as the most specific skill in algebra word problem solution, and problem type identification as the most general; the latter being explained quite well by g/working memory and verbal knowledge.

Predicting PASCAL Tutorial Performance

Continuing in an exploratory mode, we tried to determine how well the complete set of factors accounted for performance on the PASCAL tutorial. The previous analysis indicated considerable overlap between problem type identification and the working memory and verbal knowledge factors, and modest overlap between problem decomposition and sequencing and reasoning. We asked next whether the word problem solving components would still have significant contributions in predicting PASCAL tutorial performance with the general cognitive factors in the model.

The analysis proceeded as before, with the exception that we introduced two new variables as indicators of specific prior knowledge of computer programming. These were SPKNOW, the score on the pretest of computer programming concepts, and PRGCRS, a binary vector coding past coursework in non-PASCAL computer programming.[3] About 43% of the variance in SPKNOW was explained by g/working memory (loading = .46) and reasoning (loading = .47); consequently, we simply included SPKNOW as an additional indicator of these two factors. PRGCRS did not correlate significantly with other variables; thus, we included it as an independent observed variable in the path modeling.

Model 8a includes six factors and one observed variable as predictors of PASCAL tutorial performance. This model resulted in a nonsignificant *chi-square* of 407.24 (368 *df*, p = .077) and non-normed fit index of .996. Path coefficients were significant for five variables, and were nonsignificant for problem type identification (z = -.68) and verbal knowledge (z = 1.61). About 81% of the PASCAL tutorial factor variance was explained. We then evaluated Model 8b (Fig. 8) which eliminated the nonsignificant path from problem identification to the criterion. The fit indices were hardly affected (*chi-square* = 407.9, p = .079, non-normed fit index = .996), but the path from verbal knowledge to the criterion became significant (z = 3.04). This revised model accounted for 73.6% of the criterion factor variance.

Model 8b illustrates three important points. First, g/working memory is the most potent predictor of the first stage of programming skill acquisition. This finding replicates previous work in our laboratory (e.g., Kyllonen & Soule, 1988; Shute & Kyllonen, 1990; Woltz, 1988). Second, general reasoning skill had a modest but significant direct effect independent of g/working memory. Third, relatively specific knowledge and skill reflected in either related experience (PRGCRS) or word problem translation added to more general abilities in predicting PASCAL learning.

[3] Subjects who had taken a PASCAL course were eliminated from the sample, but those who had taken other programming courses (e.g., FORTRAN, BASIC, COBOL, etc.) were retained. The binary vector was coded 1 for having had one or more programming courses and 0 for none. About 52% of the remaining 130 subjects had taken a non-PASCAL programming course. No remaining subject had taken more than one programming course.

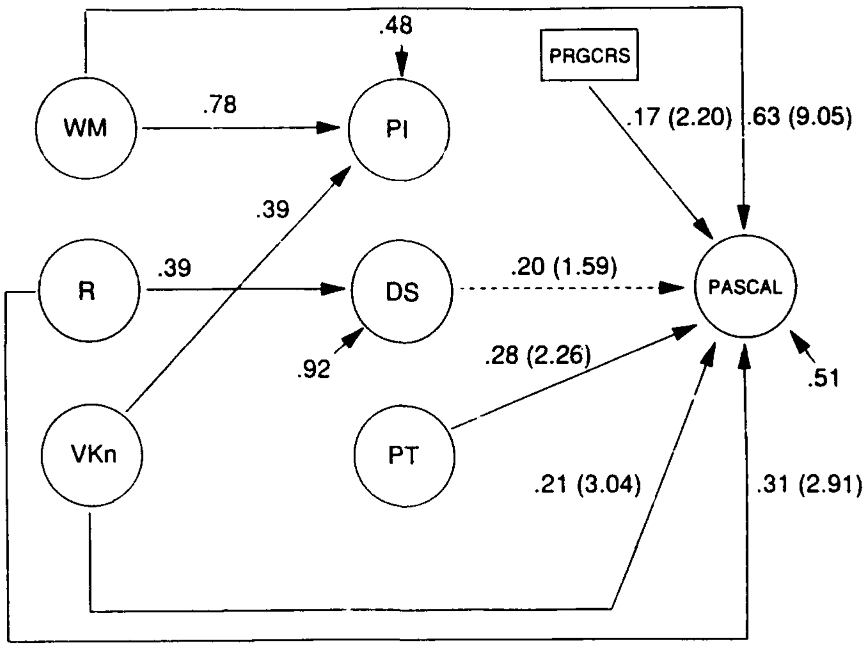

Figure 8. Path model (Model 8) relating cognitive factors to PASCAL tutorial performance. Indicators for factors (observed dependent variables) are not shown.

DISCUSSION

This study demonstrated several important points about the structure of cognitive abilities and their relationship to the initial phase of skill acquisition. First, our data suggested that certain cognitive abilities often thought to be alternative indicators of general cognitive ability should instead be considered separate abilities. We found that working memory capacity, reasoning, and verbal knowledge were distinguishable abilities and that each made significant unique contributions to the prediction of skill acquisition. This set of results is incompatible with the notion that working memory, reasoning, and verbal knowledge are simply interchangeable indicators of general cognitive ability.

Second, this study demonstrated that algebra word problem translation skill added to the more general abilities such as general reasoning, working memory, and verbal knowledge in predicting success in the first stage of programming skill acquisition. The contribution of word problem translation skill was independent of these other abilities and of previous formal instruction in another programming language.

Somewhat contrary to expectations, confirmatory factor analyses indicated that a two-factor model reflecting declarative understanding and procedural competence was not superior in fit to a one-factor model of PASCAL learning outcome. This finding is consistent with the notion that in complex skills such as programming it may take many hours of instruction and practice before declarative knowledge becomes transformed into a procedural representation.

The results of this study suggest a minor modification of Ackerman's theory (Ackerman, 1988) of how cognitive abilities relate to individual differences in skill acquisition. Ackerman (1988) suggested that individual differences in the initial declarative stage of skill learning were mostly due to variations in general ability, which he equated with the amount or efficiency of attentional resources or working memory (Ackerman, 1987). Ackerman (1988) also suggested that if the skill to be learned was primarily verbal, spatial, or quantitative, then group factors representing these modalities might also be predictive of skill acquisition. Our findings suggest that relatively specific reasoning skills that are similar if not isomorphic to the new skill domain will transfer, i.e., will add to more general abilities in predicting initial declarative learning of the skill.

In general, our results are consistent with the theoretical perspective of Mayer (1985; 1987; Mayer et al., 1986). In our study each of the components found to underlie algebra word problem solution contributed to the prediction of programming skill learning when considered by themselves. In Mayer's theory these components represent the existing conceptual knowledge that the students can transfer to the new skill. Our classification is somewhat different. We view the problem type identification component as primarily conceptual knowledge, similar to what others have called categorization skills in studies of experts and novices in other problem solving domains (e.g., Chi, Feltovich, & Glaser, 1981). However, our data suggest that the level of conceptual knowledge a student achieves is dependent on his working memory capacity and perhaps to a lesser extent, on the amount of general knowledge which can be used to assimiliate new knowledge.

On the other hand, we view the decomposition and sequencing and the problem translation components as primarily procedural knowledge, perhaps specialized versions of the general reasoning skills tapped by the Culture Fair Test. This distinction between conceptual and procedural knowledge may not be as important as the possibility that relatively specific skills transfer to new domains. The contribution of word problem translation skill to the prediction of learning PASCAL was significant even when the effects of working memory, general verbal knowledge, and general reasoning ability were statistically controlled. Future research should attempt replication of these findings with a longer, more extensive observation of the skill acquisition process that includes the development of procedural skill.[4]

The present study has practical implications for selection and classification testing in the armed services. For certain technical specialties such as computer programming, tests that assess cognitive skills analogous to those that must be developed in that technical specialty might have significant incremental predictive validity. It may prove beneficial to conduct cognitive task analyses for some technical specialties to identify skills not currently assessed by the standard test battery, i.e., the Armed Services Vocational Aptitude Battery. Future applied research might address the feasibility of this approach.

[4]Shute & Kyllonen (1990) described research at our lab in which various cognitive tests were administered to students learning PASCAL from an Intelligent Tutoring System (ITS). The average completion time was 12.3 hours. Correlations between learning outcome and variables similar to those reported here were as follows: .56 to .60 for working memory, .54 for general verbal knowledge, and .60 for specific prior knowledge. Correlations were generally higher for the Shute study and may reflect the effects of either (1) a wider range of talent, or (2) greater skill differentiation over time.

REFERENCES

Ackerman, P.L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3-27.

Ackerman, P.L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 288-318.

Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369-406.

Anderson, J.R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94*, 192-210.

Anderson, R.C. & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews.* Newark, Delaware: International Reading Association.

Anderson, R.C., & Pearson, P. D. (1984). A schema theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255-291). New York: Longman.

Bayman, P. (1983). *Effects of instructional procedure on learning a first programming language.* Unpublished doctoral dissertation. University of California, Santa Barbara.

Bentler, P.M. (1989). *EQS: Structural equations program manual.* Los Angeles, CA: BMDP Statistical Software.

Bjorklund, D.F. (1987). How age changes in knowledge base contribute to the development of children's memory: An interpretive review. *Developmental Review, 7*, 93-130.

Brooks, R. (1977). Toward a theory of the cognitive processes in computer programming. *International Journal of Man-Machine Studies, 9*, 737-751.

Cattell, R.B. (1971). *Abilities: Their structure, growth, and action.* Boston, MA: Houghton-Mifflin.

Cattell, R.B. (1973). *Technical supplement for the Culture Fair Intelligences Tests - Scales 2 and 3.* Champaign, IL: Institute for Personality and Ability Testing.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Hunt, E., & Lansman, M. (1982). Individual differences in secondary task performance. *Memory and Cognition, 10*(1), 10-24.

Kenney, D. A. (1979). *Correlation and causality.* New York: Wiley.

Kyllonen, P.C., & Christal, R.E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence, 14*, 389-433.

Kyllonen, P.C., & Soule, C. (1988, April). *The role of working memory and general problem solving skill in acquiring computer programming skill.* Paper presented at the American Educational Research Association, New Orleans, LA.

Kyllonen, P.C., & Tirre, W.C. (1988). Individual differences in associative learning and forgetting. *Intelligence, 12,* 393-421.

Kyllonen, P.C., Tirre, W.C., & Christal, R.E. (1989). *Knowledge and processing speed as determinants of associative learning* (AFHRL-TP-87-68, AD-A207 106). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Kyllonen, P.C., & Woltz, D.J. (1989). *The role of cognitive factors in the acquisition of cognitive skill* (AFHRL-TP-89-5, AD-A218 069). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Mayer, R.E. (1985). Learning in complex domains: A cognitive analysis of computer programming. *Psychology of Learning and Motivation, 19,* 89-130.

Mayer, R.E. (1987). Cognitive aspects of learning and using a programming language (pp. 61-79). In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human-computer interaction.* Cambridge, MA: MIT Press.

Mayer, R.E., Dyck, J.L., & Vilberg, W. (1986). Learning to program and learning to think: What's the connection? *Communications of the ACM, 29* (7), 605-610.

Pena, M.C. (1989). *Cognitive determinants of the early stages of programming skill acquisition.* Unpublished master's thesis, St. Mary's University.

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods and Research, 13,* 109-119.

Shute, V.J., & Kyllonen, P.C. (1990). *Modeling individual differences in programming skill acquisition* (AFHRL-TP-90-76, AD-A229 816). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Shute, V.J., & Pena, M.C. (1990). *Acquisition of programming skills.* (AFHRL-TP-89-54, AD-A221 036). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Tirre, W.C. (1989). *Knowledge and ability factors underlying simple learning by accretion.* (AFHRL-TP-89-48, AD-A214 450). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

U.S. Department of Defense (1984). *Armed Services Vocational Aptitude Battery (ASVAB): Test manual.* North Chicago, IL: U.S. Military Entrance Processing Command.

Woltz, D.J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General, 117*(3), 319-331.

APPENDIX

Descriptive Statistics and Internal Consistency Estimates of Experimental Tests

Experimental Test	*Mean*	*SD*	$R_{xx'}$
PASCAL Tutorial			
Section 1	68.6	19.2	.27
Section 2	70.4	22.5	.55
Section 3	50.7	27.1	.95
Section 4	58.0	16.7	.32
Section 5	72.8	13.6	.56
Section 6	66.4	13.7	.35
Composite Reliability			.84
Working Memory Capacity			
Continuous Paired Associates	68.8	14.4	.93
Alphabet Recoding	48.4	18.7	.87
Backward Digit Span	50.1	18.6	.85
Composite Reliability			.94
Culture Fair Intelligence Test			
CFIT1: Series	54.7	19.1	.63
CFIT2: Classifications	52.9	13.6	.59
CFIT3: Matrices	45.7	13.1	.22
CFIT4: Conditions	46.8	24.5	.67
Composite Reliability			.73
Algebra Problem Solving Test			
Problem Identification	57.1	19.0	.69
Problem Decomposition and Sequencing	43.7	25.3	.73
Problem Translation	51.5	23.5	.78
Computer Programming Knowledge			
Programming Concepts 1	57.6	9.7	.33
Programming Concepts 2	67.7	12.3	.31
Composite Reliability			.37

Note. N = 130. Tests are described in text. All scores are percent correct accuracy.